

# Le experimentos con le machina OPUS CAT

Bohdan Smilauer, Praga, b.smilauer@post.cz

6e Auguste 2021:

Le theoria vide

[https://aclanthology.org/2021\\_eacl-demos.34.pdf](https://aclanthology.org/2021_eacl-demos.34.pdf)

SW que esseva usate in iste experimento

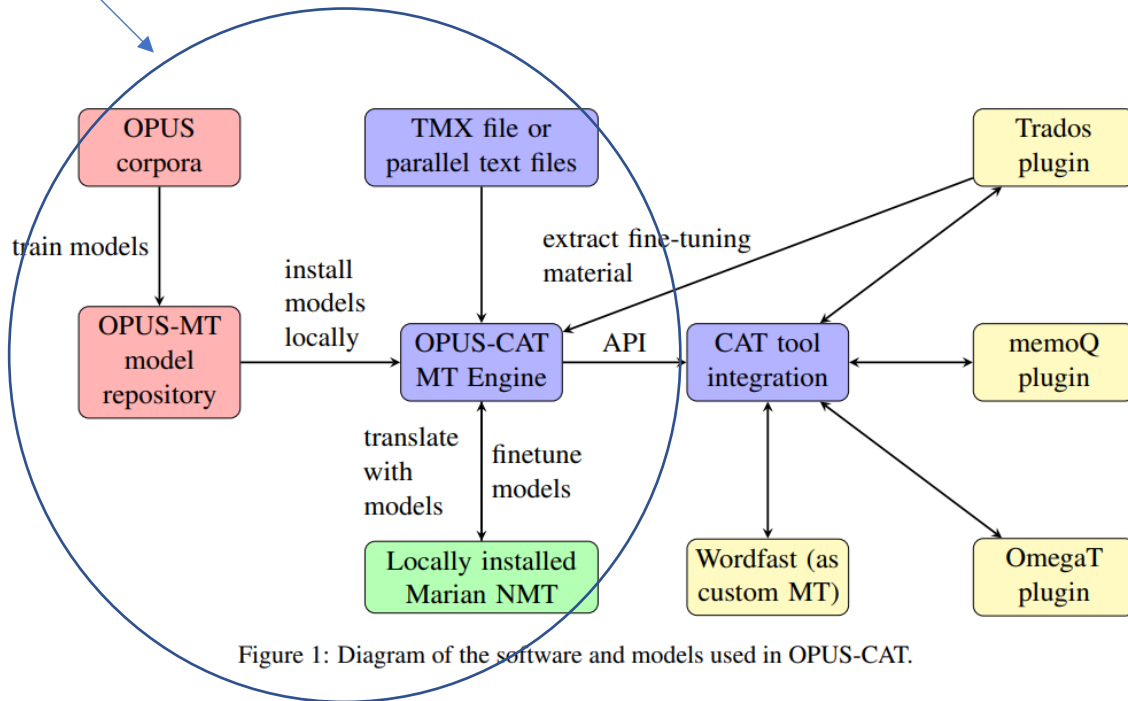


Figure 1: Diagram of the software and models used in OPUS-CAT.

Installation del OPUS/CAT MT Machine + Locally installed Marian NMT

Guida pro le installation:

<https://helsinki-nlp.github.io/OPUS-CAT/>

https://helsinki-nlp.github.io/OPUS-CAT/install

## Quickstart for translators

1. Download the [OPUS-CAT MT Engine](#) and install it to your computer by extracting it in a folder on your computer. **IMPORTANT:** OPUS-CAT MT Engine generates the machine translation, and all OPUS-CAT plugins require that the OPUS-CAT MT Engine is installed on the same computer and running when the plugins are used.

Discarga:

[https://github.com/Helsinki-NLP/OPUS-CAT/releases/download/engine\\_v1.1.0.4/OpusCatMtEngine\\_v1.1.0.4.zip](https://github.com/Helsinki-NLP/OPUS-CAT/releases/download/engine_v1.1.0.4/OpusCatMtEngine_v1.1.0.4.zip)

Post le discargation 39 MB

<input checked="" type="checkbox"/>	OpusCatMtEngine_v1.1.0.4.zip	07/08/2021 00:47	Komprimovaná složk...	39,152 kB
-------------------------------------	------------------------------	------------------	-----------------------	-----------

Extrahe le programmas de .zip

OpusCatMtEngine\_v1.1.0.4

Sdílení Zobrazení

Tento počítač > Stažené soubory > OpusCatMtEngine\_v1.1.0.4 >

Název	Datum změny	Typ	Velikost
Evaluation	07/08/2021 00:49	Složka souborů	
Marian	07/08/2021 00:49	Složka souborů	
Preprocessing	07/08/2021 00:49	Složka souborů	
TatoebaTestsets	07/08/2021 00:53	Složka souborů	
x64	07/08/2021 00:54	Složka souborů	
x86	07/08/2021 00:54	Složka souborů	
Autofac.dll	07/08/2021 00:49	Rozšířená aplikace	312 kB
Autofac.Integration.Owin.dll	07/08/2021 00:49	Rozšířená aplikace	17 kB
Autofac.Integration.Owin.pdb	07/08/2021 00:49	Program Debug Data...	13 kB
Autofac.Integration.Owin.xml	07/08/2021 00:49	Dokument ve formát...	22 kB
Autofac.Integration.WebApi.dll	07/08/2021 00:49	Rozšířená aplikace	52 kB
Autofac.Integration.WebApi.Owin.dll	07/08/2021 00:49	Rozšířená aplikace	8 kB
Autofac.Integration.WebApi.Owin.pdb	07/08/2021 00:49	Program Debug Data...	12 kB
Autofac.Integration.WebApi.Owin.xml	07/08/2021 00:49	Dokument ve formát...	2 kB
Autofac.Integration.WebApi.pdb	07/08/2021 00:49	Program Debug Data...	64 kB
Autofac.Integration.WebApi.xml	07/08/2021 00:49	Dokument ve formát...	91 kB
Autofac.pdb	07/08/2021 00:49	Program Debug Data...	410 kB
Autofac.xml	07/08/2021 00:49	Dokument ve formát...	705 kB
customize.yml	07/08/2021 00:49	Soubor YML	1 kB
EntityFramework.dll	07/08/2021 00:49	Rozšířená aplikace	4,872 kB

continuation:

Soubory > OpusCatMtEngine\_v1.1.0.4

Název	Datum změny	Typ	Velikost
customize.yml	07/08/2021 00:49	Soubor YML	1 kB
EntityFramework.dll	07/08/2021 00:49	Rozšířená aplikace	4,872 kB
EntityFramework.SqlServer.dll	07/08/2021 00:49	Rozšířená aplikace	578 kB
EntityFramework.SqlServer.xml	07/08/2021 00:49	Dokument ve formát...	160 kB
EntityFramework.xml	07/08/2021 00:49	Dokument ve formát...	3,650 kB
Evaluate.bat	07/08/2021 00:49	Dávkový soubor syst...	1 kB
LiveCharts.dll	07/08/2021 00:49	Rozšířená aplikace	149 kB
LiveCharts.pdb	07/08/2021 00:49	Program Debug Data...	356 kB
LiveCharts.Wpf.dll	07/08/2021 00:49	Rozšířená aplikace	213 kB
LiveCharts.Wpf.pdb	07/08/2021 00:49	Program Debug Data...	488 kB
LiveCharts.Wpf.xml	07/08/2021 00:49	Dokument ve formát...	172 kB
LiveCharts.xml	07/08/2021 00:49	Dokument ve formát...	216 kB
Microsoft.Bcl.AsyncInterfaces.dll	07/08/2021 00:49	Rozšířená aplikace	21 kB
Microsoft.Bcl.AsyncInterfaces.xml	07/08/2021 00:49	Dokument ve formát...	18 kB
Microsoft.Owin.dll	07/08/2021 00:49	Rozšířená aplikace	99 kB
Microsoft.Owin.Host.HttpListener.dll	07/08/2021 00:49	Rozšířená aplikace	84 kB
Microsoft.Owin.Host.HttpListener.xml	07/08/2021 00:49	Dokument ve formát...	8 kB
Microsoft.Owin.Hosting.dll	07/08/2021 00:49	Rozšířená aplikace	65 kB
Microsoft.Owin.Hosting.xml	07/08/2021 00:49	Dokument ve formát...	49 kB
Microsoft.Owin.xml	07/08/2021 00:49	Dokument ve formát...	168 kB

Stahované soubory > OpusCatMtEngine\_v1.1.0.4

Curre iste program pro le installation

Název	Datum změny	Typ	Velikost
Microsoft.Owin.xml	07/08/2021 00:49	Dokument ve formát...	168 kB
Newtonsoft.Json.dll	07/08/2021 00:49	Rozšíření aplikace	491 kB
Newtonsoft.Json.xml	07/08/2021 00:49	Dokument ve formát...	468 kB
OpusCatMTEngine.exe	07/08/2021 00:49	Aplikace	1,114 kB
OpusCatMTEngine.exe.config	07/08/2021 00:49	XML Configuration File	6 kB
OpusCatMTEngine.pdb	07/08/2021 00:49	Program Debug Data...	530 kB
OpusMTInterface.dll	07/08/2021 00:49	Rozšíření aplikace	7 kB
OpusMTInterface.pdb	07/08/2021 00:49	Program Debug Data...	12 kB
Owin.dll	07/08/2021 00:49	Rozšíření aplikace	5 kB
Serilog.dll	07/08/2021 00:49	Rozšíření aplikace	124 kB
Serilog.Sinks.File.dll	07/08/2021 00:49	Rozšíření aplikace	28 kB
Serilog.Sinks.File.pdb	07/08/2021 00:49	Program Debug Data...	7 kB
Serilog.Sinks.File.xml	07/08/2021 00:49	Dokument ve formát...	33 kB
Serilog.xml	07/08/2021 00:49	Dokument ve formát...	272 kB
StartMosesBpeMtPipe.bat	07/08/2021 00:49	Dávkový soubor syst...	1 kB
StartMosesBpeMultilingualMtPipe.bat	07/08/2021 00:49	Dávkový soubor syst...	1 kB
StartSentencePieceMtPipe.bat	07/08/2021 00:49	Dávkový soubor syst...	1 kB
StartSentencePieceMultilingualMtPipe.bat	07/08/2021 00:49	Dávkový soubor syst...	1 kB
System Buffers.dll	07/08/2021 00:49	Rozšíření aplikace	21 kB

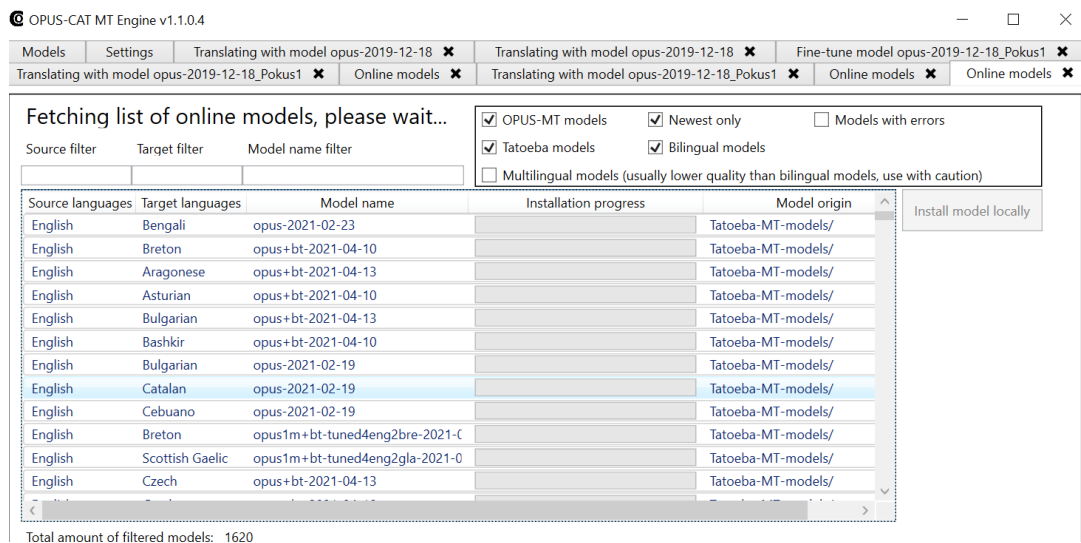
Stahované soubory > OpusCatMtEngine\_v1.1.0.4

Název	Datum změny	Typ	Velikost
StartSentencePieceMultilingualMtPipe.bat	07/08/2021 00:49	Dávkový soubor syst...	1 kB
System Buffers.dll	07/08/2021 00:49	Rozšíření aplikace	21 kB
System Buffers.xml	07/08/2021 00:49	Dokument ve formát...	4 kB
System.Data.SQLite.dll	07/08/2021 00:50	Rozšíření aplikace	356 kB
System.Data.SQLite.dll.config	07/08/2021 00:50	XML Configuration File	1 kB
System.Data.SQLite.EF6.dll	07/08/2021 00:50	Rozšíření aplikace	183 kB
System.Data.SQLite.Linq.dll	07/08/2021 00:50	Rozšíření aplikace	183 kB
System.Data.SQLite.xml	07/08/2021 00:50	Dokument ve formát...	1,077 kB
System.Diagnostics.DiagnosticSource.dll	07/08/2021 00:50	Rozšíření aplikace	65 kB
System.Diagnostics.DiagnosticSource.xml	07/08/2021 00:50	Dokument ve formát...	34 kB
System.Memory.dll	07/08/2021 00:50	Rozšíření aplikace	138 kB
System.Memory.xml	07/08/2021 00:50	Dokument ve formát...	14 kB
System.Net.Http.Formatting.dll	07/08/2021 00:50	Rozšíření aplikace	175 kB
System.Net.Http.Formatting.xml	07/08/2021 00:50	Dokument ve formát...	189 kB
System.Numerics.Vectors.dll	07/08/2021 00:50	Rozšíření aplikace	114 kB
System.Numerics.Vectors.xml	07/08/2021 00:50	Dokument ve formát...	180 kB
System.Runtime.CompilerServices.Unsafe.dll	07/08/2021 00:50	Rozšíření aplikace	17 kB
System.Runtime.CompilerServices.Unsafe.xml	07/08/2021 00:50	Dokument ve formát...	14 kB
System.Threading.Tasks.Extensions.dll	07/08/2021 00:50	Rozšíření aplikace	33 kB
System.Threading.Tasks.Extensions.xml	07/08/2021 00:50	Dokument ve formát...	10 kB
System.Runtime.CompilerServices.Unsafe.xml	07/08/2021 00:50	Dokument ve formát...	14 kB
System.Threading.Tasks.Extensions.dll	07/08/2021 00:50	Rozšíření aplikace	33 kB
System.Threading.Tasks.Extensions.xml	07/08/2021 00:50	Dokument ve formát...	10 kB
System.Web.Http.dll	07/08/2021 00:50	Rozšíření aplikace	445 kB
System.Web.Http.Owin.dll	07/08/2021 00:50	Rozšíření aplikace	61 kB
System.Web.Http.Owin.xml	07/08/2021 00:50	Dokument ve formát...	16 kB
System.Web.Http.xml	07/08/2021 00:50	Dokument ve formát...	528 kB
ThirdPartyNotices.txt	07/08/2021 00:53	Textový dokument	41 kB
TranslateBatchSentencePiece.bat	07/08/2021 00:53	Dávkový soubor syst...	1 kB
ValidateBpe.bat	07/08/2021 00:53	Dávkový soubor syst...	3 kB
ValidateSp.bat	07/08/2021 00:54	Dávkový soubor syst...	3 kB
YamlDotNet.dll	07/08/2021 00:54	Rozšíření aplikace	196 kB
YamlDotNet.xml	07/08/2021 00:54	Dokument ve formát...	253 kB

Post le installation le dialog se appere

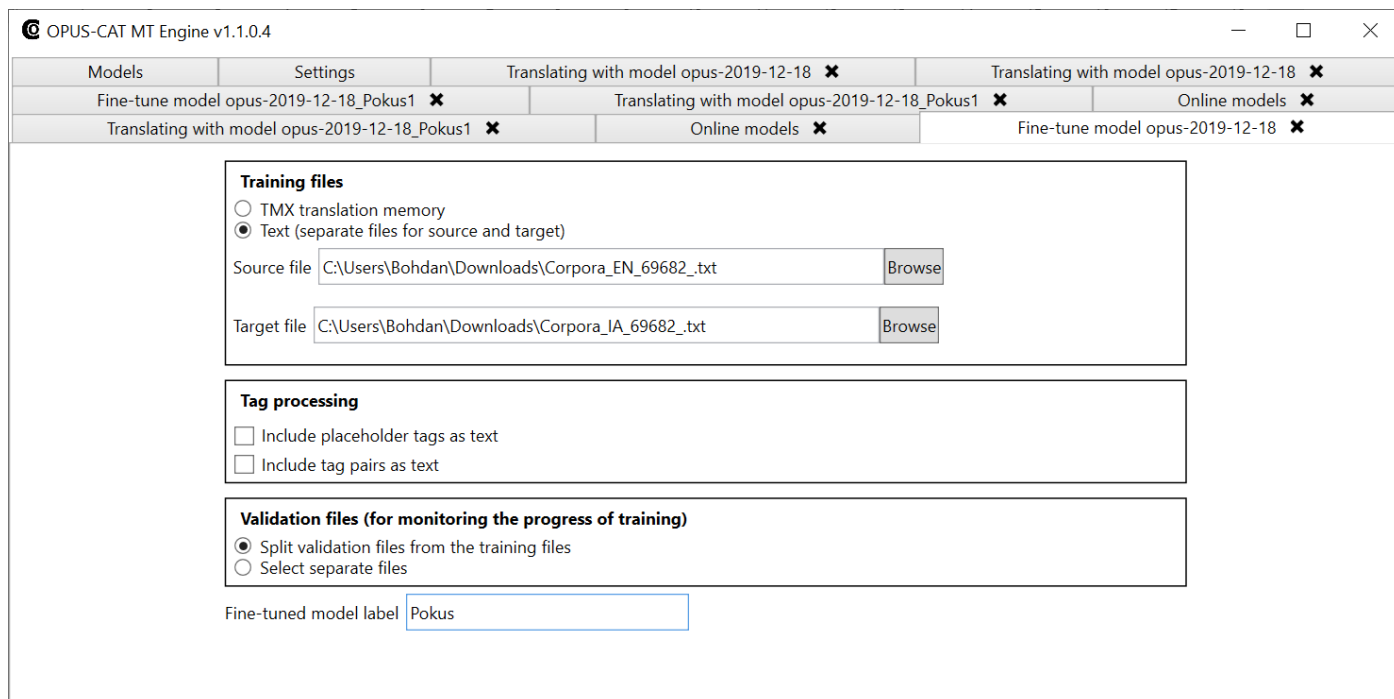
## Models

### Install OPUS Modules from Web



### Installation del English - Catalan

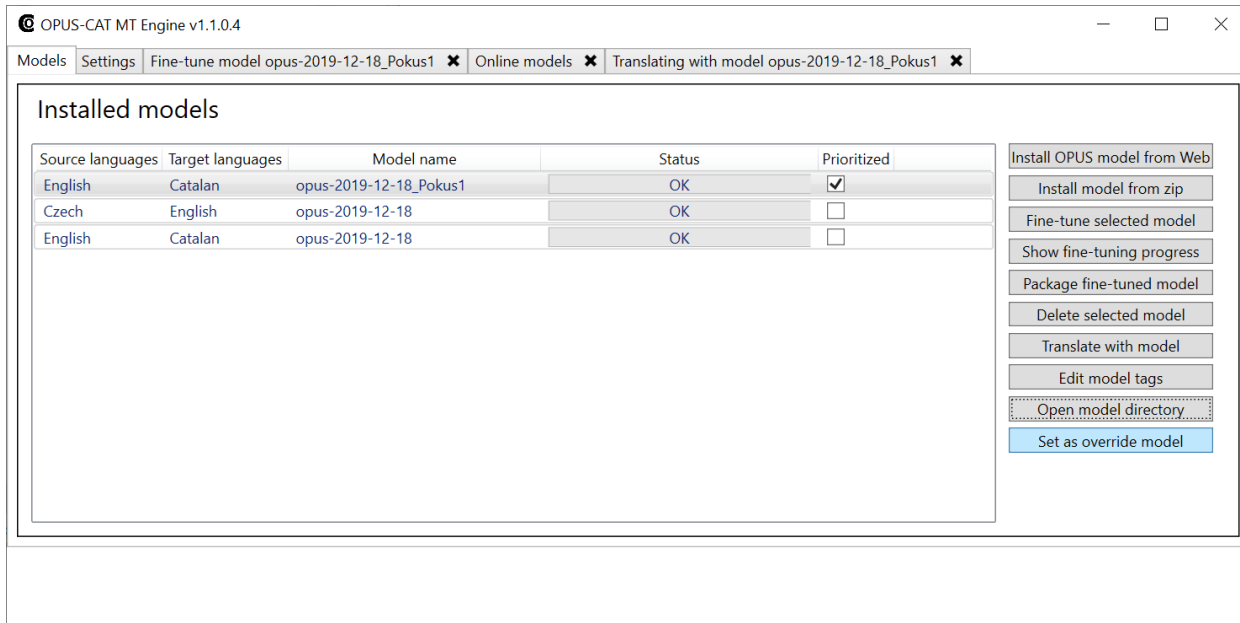
#### Fine tune selected model



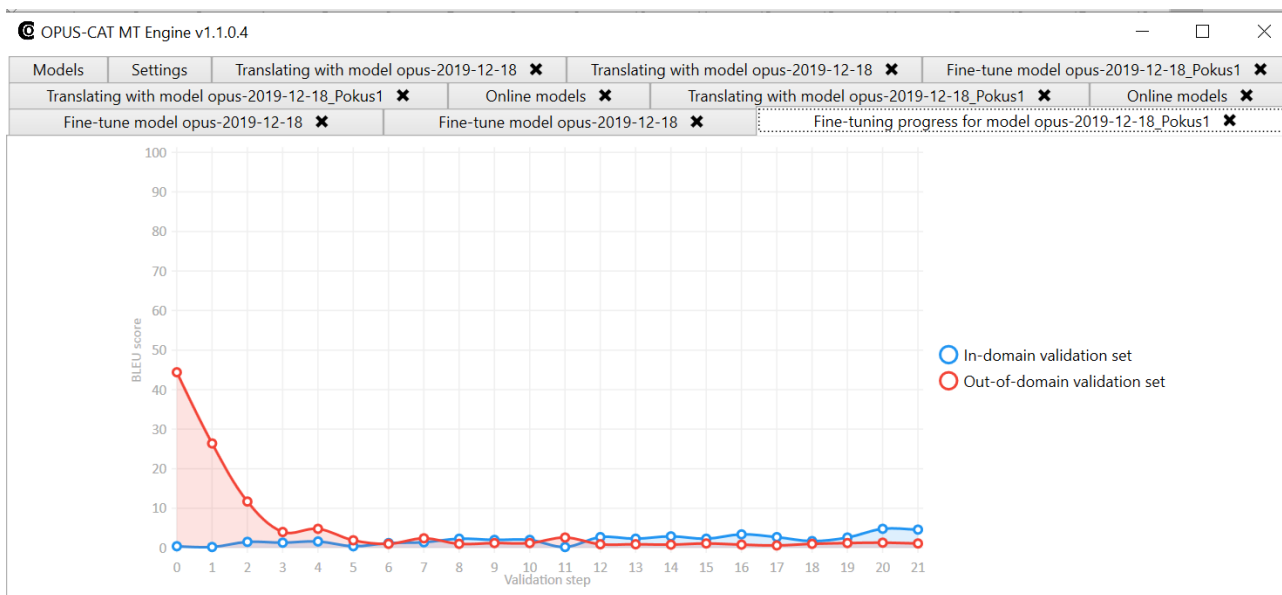
Le filios [Corpora\\_EN\\_69682.txt](http://www.interlingva.cz/Corpora_EN_69682.txt), Le filios [Corpora\\_IA\\_69682.txt](http://www.interlingva.cz/Corpora_IA_69682.txt), son ganiate como le listas del phrases del corpora

[http://www.interlingva.cz/Corpora\\_IA\\_CZ\\_EN\\_69682.xlsx](http://www.interlingva.cz/Corpora_IA_CZ_EN_69682.xlsx) in Excel

Clica Fine-tune selected model, le processo durava circa 5 horas.



Set override model



## le fascie del OPUS MT

-2019-12-18\_Pokus1

dílení Zobrazení

Bohdan > AppData > Local > opuscat > models > en-ca > opus-2019-12-18\_Pokus1

	Název	Datum změny	Typ	Velikost
Image	combined.source	07/08/2021 01:39	Soubor SOURCE	11 kB
	combined.target	07/08/2021 01:39	Soubor TARGET	11 kB
	custom.source	07/08/2021 01:37	Soubor SOURCE	5,452 kB
	custom.target	07/08/2021 01:37	Soubor TARGET	5,775 kB
Help	customize.yml	07/08/2021 01:40	Soubor YML	2 kB
Image	decoder.yml	18/12/2019 22:16	Soubor YML	1 kB
Image	LICENSE	18/12/2019 22:16	Soubor	19 kB
	modelconfig.yml	07/08/2021 06:34	Soubor YML	1 kB
es	opus.spm32k-spm32k.transformer-align.mo...	07/08/2021 06:34	Soubor NPZ	283,192 kB
	opus.spm32k-spm32k.transformer-align.mo...	07/08/2021 06:34	Soubor YML	1 kB
	opus.spm32k-spm32k.transformer-align.mo...	07/08/2021 06:34	Soubor NPZ	566,281 kB
	opus.spm32k-spm32k.transformer-align.mo...	07/08/2021 06:34	Soubor YML	1 kB
model_packages	opus.spm32k-spm32k.transformer-align.mo...	07/08/2021 06:34	Soubor YML	5 kB
	opus.spm32k-spm32k.transformer-align.trai...	08/12/2019 13:05	Textový dokument	36 kB
	opus.spm32k-spm32k.transformer-align.vali...	08/12/2019 13:05	Textový dokument	2 kB
	opus.spm32k-spm32k.vocab.yml	08/12/2019 07:54	Soubor YML	914 kB
	postprocess.sh	18/12/2019 22:16	Shell Script	1 kB
	preprocess.sh	18/12/2019 22:16	Shell Script	1 kB
2019-12-18	preprocessed_combined.source	07/08/2021 01:39	Soubor SOURCE	11 kB
2019-12-18_Pokus1	preprocessed_combined.target	07/08/2021 01:39	Soubor TARGET	11 kB

## le examples del traduction machinal:

OPUS-CAT MT Engine v1.1.0.4

Models Settings Fine-tune model opus-2019-12-18\_Pokus1 Online models Translating with model opus-2019-12-18\_Pokus1

Note: This translation functionality is intended mainly for testing models. Sentences of the source text should be on separate lines.

Source language: English Target language: Catalan  Show subword segmentation

Source text:

An old man lived in the village.

He was one of the most unfortunate people in the world.

The whole village was tired of him;

he was always gloomy,

Alignments are available for this model (hover over word to see its aligned words, if any)

Translation:

Un vetulo viveva in le populo.

Ille esseva un del gente multo frigide in le mundo.

Le marionette esseva plen de iste;

ille esseva sempre triste,

EN original	IA Traduction machinal
An old man lived in the village.	Un vetulo viveva in le populo.
He was one of the most unfortunate people in the world.	Ille esseva un del gente multo frigide in le mundo.
The whole village was tired of him;	Le marionette esseva plen de iste;
he was always gloomy,	ille esseva sempre triste,
he constantly complained and was always in a bad mood.	ille se demandava multo, e sempre esseva in mal humor.
The longer he lived, the more bile he was becoming and the more poisonous were his words.	Le prime ille viveva, le plus bile que esseva plus vermente esseva su parolas.
People avoided him, because his misfortune became contagious.	Le gente le attendeva, proque su infortunatia esseva contagie.
It was even unnatural and insulting to be happy next to him.	Isto esseva ascoltate e attendeva a ser felice al parte de ille.
He created the feeling of unhappiness in others	Ille credeva le sentimento de impervenibilitate in le altere
But one day, when he turned eighty years old, an incredible thing happened.	Ma un die, quando ille haveva vidite 80 annos, un cosas eccellente haveva succedite.
Instantly everyone started hearing the rumour:"	Isto non poteva audir le rumorulo:
An Old Man is happy today, he doesn't complain about anything, smiles, and even his face is freshened up.	Un vetule homine es felice, ille non demandan de reso, ride e su facie se ascoltava.
The whole village gathered together.	Tote le populo se reuniva.
The old man was asked:	Le vetulo diceva:
Villager: What happened to you?	Qual cosa te haveva attendite?
Nothing special.	Non special.
Eighty years I've been chasing happiness, and it was useless.	Iste annos io haveva cercate le felicitate, e illo esseva servite.
And then I decided to live without happiness and just enjoy life.	E io debeva vider sin felicitate e solmente gaudia le vita.
That's why I'm happy now." -	Isto es proque io es felice.
An old man lived in the village.	Un vetulo viveva in le populo.
He was one of the most unfortunate people in the world.	Ille esseva un del gente multo frigide in le mundo.

Le fin